

Latest notable achievements in genomics

WU JiaYan[†], XIAO JingFa[†] & YU Jun^{*}

Key Laboratory of Genomics Science and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China

Received May 9, 2012; accepted May 22, 2012

Citation: Wu J Y, Xiao J F, Yu J. Latest notable achievements in genomics. *Sci China Life Sci*, 2012, 55: 645–648, doi: 10.1007/s11427-012-4331-x

Within the past 10 years, the progress of genomics development in China exhibited a frog-leap in terms of scale, quality, organization, and international collaboration. Human genome study has entered a new phase of interdisciplinarity in the history of life science in China. The rise of genomics relied on the development of technology and its integration with other academic disciplines, particularly in the following three areas. First, DNA sequencing technology has been improving its productivity and efficiency at an exponential growth rate. Second, bioinformatics supports the system of data collection, administration, annotation, distribution, and analysis for genome researchers. Third, the integration with the genomics data and other “omics” data provides hypotheses and building models for systems biology research. In 2009 and 2010, *Science China Life Sciences* published over 20 research papers and review articles about genomics. Authors of those publications are all leading life scientists, shared their understanding on different hot questions or focused on a single question of common interest from different views arisen from their respective expertise.

Technological advancement in DNA sequencing has transformed the study of fundamental elements of life—from individual, localized or clustered genes to the whole genomes, which in turn demands more competent and efficient sequencing technology. Zhou XiaoGuang *et al.* summarized all major emerging sequencing platforms [1–3]

and provided a review of sequencing technology evolution, generational advancements with their merits and drawbacks, and prediction of possible direction of the field [4]. With the second generation platforms in hands and the third or single molecule platforms in the horizon, the goal of reaching “one thousand dollar per genome” will soon become tangible. Aside from sequencing technology, other important tools for large-scale and high-throughput genome research, such as various DNA microarrays and genotyping tools, are also brought out for revealing both structural and functional characteristics of genomes and transcriptomes [5]. We expect that Chinese scientists not only are among the best to utilize the emerging new technologies for genomics research but also join the effort in developing alternatives to create friendly competitions that will ensure rapid and less expensive readout of sequence information providing ample possibilities in comparative genomics, disease diagnosis, and ultimately personalized (or individualized) medicine.

In China, the incidence of common diseases has been increasing, such as cancer and cardiovascular diseases. Newly emerging infectious agents are also a major threat to human health. In addition, there is a persistently high rate of hepatitis B virus (HBC) infection resulting in a large number of chronic hepatitis patients and some of them have a strong tendency to develop liver cirrhosis and eventually liver cancer. Collaborative research activities in this front have been quite fruitful [6]. Excellent investigations have been performed in etiology-associated genomic and genetic researches [7,8] as well as genomic instability and

[†]Contributed equally to this work

^{*}Corresponding author (email: junyu@big.ac.cn)

expression pattern analysis of liver cancers [9–11]. Huang Jian has reviewed some recent epigenetics research progress referring to heritable changes for hepatocarcinogenesis, including DNA methylation, genome imprinting, chromatin remodeling, histone modification, and microRNA regulations [12]. As to the research area of pathogenic bacteria, rapid accumulation of bacterial genomic data offered unprecedented opportunities to understand bacterial biology from a holistic point of view [13]. Recent progress in pathogenic bacterial genomics in China encompasses genomic diversity, molecular epidemiology, microbial forensics [14–16], diagnostics [17], vaccine design [18–20], novel antibacterial agent development [21], and metagenomics for comprehensive understanding of human health [22].

Chinese genomicists have been working in the frontier of plant and animal genomics, such as rice [23] and giant panda [24]. The former not only produces the first plant genome sequence using the whole-genome shotgun strategy, but also is still generating large amounts of sequence information in genetics and gene expression. Li LiHua *et al.* [25] investigated the gene expression profile in rice roots at 6, 24 and 72 h under low phosphorus (P) stress and compared it with a control (normal P) profile, providing useful information in the molecular process associated with Pi deficiency and facilitating research on improving Pi utilization in crop species. Kou YanJun *et al.* [26] analyzed the colocalization of a set of characterized rice defense-responsive genes and resistance QTLs against different pathogens and the expression patterns of these genes, which suggests that defense-responsive genes are important resources of resistance QTLs in rice. As to the giant panda (*Ailuropoda melanoleuca*), one of the most endangered species in the world known as the “Chinese treasure”, complete Sanger sequencing, assembly, annotation and comparative analysis were carried out on the selected BACs of a joined length of 878 kb, which provided detailed sequence and structure information for new genes and repeats of the giant panda for further studies [27].

At the mean time, investigations on genomics of indigenous species and important genes are still hotspots in China. The study on Tibetan chicken's embryonic adaptability to chronic hypoxia showed that Tibetan chickens have much higher hatch-ability than low-land chicken breeds in high-altitude areas of Tibet [28]. Most *Sox* genes directly affect cell fate determination and differentiation. Lin YuShuang *et al.* [29] isolated two *Sox* genes from the closet living invertebrate relative of the vertebrates, *Branchiostoma belcheri*, and aligned their protein sequences, discovering that vertebrate homologs might arise from gene duplications during evolution. Miao YongWang *et al.* [30] figured out the allele E^{BR} or the allele carrying the amino acid p104S in melanocortin-1 receptor (MC1R) is associated

with the black coat color in buffalo, based on their genotype and phenotype analysis of the hybrid offspring of black river buffalo and gray swamp buffalo. Revealed by the full-length B' cDNA and its encoding protein, Fu Hu *et al.* [31] suggested that B' appears to play a very important role during gold fish (*Carassius auratus*) development and also in adult tissue homeostasis. B' protein is strongly expressed at the stage of two-cell, multiple-cell, blastula, gastrula, neurula, and optic vesicle, and then decreased at the stages of brain differentiation and eye pigmentation. He Jing *et al.* [32] observed episodic evolution in the pre hormone gene sequences of the two members of a structure-function-related hormone family, *Motilin* and *ghrelin*, during primitive placental mammal evolution, and predict that some of the lineage-specific physiological adaptations are due to episodic evolution of these two genes.

The giant leap in sequencing technologies have sent the cost of sequencing tumbling, and the daily output of sequence has soared just in the past few years [33]. Although computational biology cannot handle all demands created by the exponentially increasing amounts of sequence data, bioinformatics and biostatistics have been recognized seriously by increasingly more experimental life scientists. Chinese bioinformatic researchers have done great jobs in understanding structural dynamics of genomes. Transcriptional regulation is one of the most critical pipelines in biological process, in which *cis*-elements work as gene expression regulators. By means of analyzing activities or behaviors of the shared *cis*-elements, Hao Pei *et al.* [34] discovered those elements participating in controlling the expression pattern of “head-to-head” gene pairs and the whole transcription regulation system. Zheng LingLing and Qu LiangHu reviewed the main issues in “Computational RNAomics” of data storage and management, ncRNA gene identification and characterization, ncRNA target identification and functional prediction, and summarized the main methods [35]. Explaining the “Tragedy of the Commons” of the evolution of cooperation remains one of the greatest problems for both biology and social science. Wang RuiWu and Shi Lei built a model and carried out simulations to show that the local but not the global stability of a cooperative interaction can be maintained if the utilization of common resource remains unsaturated, achieved by density-dependent restraint or competition among the cooperative actors [36].

Human genetic variation study was ranked as one of the top ten breakthroughs in 2007 by *Science* [37]. For better understanding of human genetic diseases, better genotyping technology is of essence. Shi JinXiu *et al.* [38] reviewed the development and application of genotyping technologies, including a great number of accurate, rapid, and cost-effective technologies for SNP analysis, which promote the research of not only simple Medialian diseases

but also complex genetic disorders. Another review summarized the commonly used technologies to discover both genomic structural variants and their potential impacts, highlighting array-based, PCR-based, and sequencing-based assays, as well as discussed limitations/challenges and gave advices on how to design more reliable genomics variation studies [39]. A specific case was investigated by Tan LiJun *et al.* [40]. They performed a genome-wide association study (GWAS) of wrist bone mineral density (BMD), combined with the functional relevance of *SOX6* in cartilage formation, which all support that *SOX6* gene works as a key gene for BMD variation.

In the past two years, Chinese scientists working in domestic institutes have been very fruitful in genomics studies, coincident with the fast economic development in China and the public funding for sciences in general. As long as more young talents in life sciences, domestically educated or returning from abroad, are joining newly emerging research laboratories, scientific productivity will be boosted in China for the years to come.

- 1 Fedurco M, Romieu A, Williams S, *et al.* BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res*, 2006, 34: e22
- 2 Margulies M, Egholm M, Altman W E, *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005, 437: 376–380
- 3 Macevitz S C. DNA sequencing by parallel oligonucleotide extensions. US patent 5750341. 1998
- 4 Zhou X, Ren L, Li Y, *et al.* Next-generation sequencing technology: A technology review and future perspective. *Sci China Life Sci*, 2010, 53: 44–57
- 5 Teng X, Xiao H. Perspectives of DNA microarray and next-generation DNA sequencing technologies. *Sci China Ser C-Life Sci*, 2009, 52: 7–16
- 6 Han Z. Recent progress in genomic research of liver cancer. *Sci China Life Sci*, 2009, 52: 24–30
- 7 Jia L, Wang X, Harris C. Hepatitis B virus X protein inhibits nucleotide excision repair. *Int J Cancer*, 1999, 80: 875–879
- 8 Zhang L, Cao Y, Song J, *et al.* The correlation between integration of HBV X, S, Pre-S, C gene and the expression of oncogenes/tumor suppressor genes in primary hepatocellular carcinoma. *Chin J Hepatol*, 1999, 7: 138–139
- 9 Xu X, Huang J, Xu Z, *et al.* Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. *Proc Natl Acad Sci USA*, 2001, 98: 15089–15094
- 10 Ye Q, Qin L, Forgues M, *et al.* Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med*, 2003, 9: 416–423
- 11 Wan D, Gong Y, Qin W, *et al.* Large-scale cDNA transfection screening for genes related to cancer development and progression. *Proc Natl Acad Sci USA*, 2004, 101: 15724–15729
- 12 Huang J. Current progress in epigenetic research for hepatocarcinogenesis. *Sci China Ser C-Life Sci*, 2009, 52: 31–42
- 13 Yang R, Guo X, Yang J, *et al.* Genomic research for important pathogenic bacteria in China. *Sci China Ser C-Life Sci*, 2009, 52: 50–63
- 14 Cui Y, Li Y, George O, *et al.* Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS ONE*, 2008, 3: e2652
- 15 Ren S, Fu G, Jiang X, *et al.* Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature*, 2003, 422: 888–893
- 16 Li Y, Dai E, Cui Y, *et al.* Different region analysis for genotyping *Yersinia pestis* isolates from China. *PLoS ONE*, 2008, 3: e2166
- 17 Zhou D, Han Y, Song Y, *et al.* DNA microarray analysis of genome dynamics in *Yersinia pestis*: Insights into bacterial genome microevolution and niche adaptation. *J Bacteriol*, 2004, 186: 5138–5146
- 18 Li B, Zhou D, Wang Z, *et al.* Antibody profiling in plague patients by protein microarray. *Microbes Infect*, 2008, 10: 45–51
- 19 Chen Z, Li B, Zhang J, *et al.* Quorum sensing affects virulence-associated proteins F1, LcrV, KatY and pH6 etc. of *Yersinia pestis* as revealed by protein microarray-based antibody profiling. *Microbes Infect*, 2006, 8: 2501–2508
- 20 Li B, Jiang L, Song Q, *et al.* Protein microarray for profiling antibody responses to *Yersinia pestis* live vaccine. *Infect Immun*, 2005, 73: 3734–3739
- 21 Zheng H, Lu L, Wang B, *et al.* Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv. *PLoS ONE*, 2008, 3: e2375
- 22 Li M, Wang B, Zhang M, *et al.* Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci USA*, 2008, 105: 2117–2122
- 23 Yu J, Hu S, Wang J, *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science*, 2002, 296: 79–92
- 24 Li R, Fan W, Tian G, *et al.* The sequence and de novo assembly of the giant panda genome. *Nature*, 2010, 463: 311–317
- 25 Li L, Qiu X, Li X, *et al.* The expression profile of genes in rice roots under low phosphorus stress. *Sci China Ser C-Life Sci*, 2009, 52: 1055–1064
- 26 Kou Y, Li X, Xiao J, *et al.* Identification of genes contributing to quantitative disease resistance in rice. *Sci China Life Sci*, 2010, 53: 1263–1273
- 27 Zheng Y, Cai J, Li J, *et al.* Sequencing, annotation and comparative analysis of nine BACs of the giant panda (*Ailuropoda melanoleuca*). *Sci China Life Sci*, 2010, 53: 107–111
- 28 Li M, Zhao C. Study on Tibetan Chicken embryonic adaptability to chronic hypoxia by revealing differential gene expression in heart tissue. *Sci China Ser C-Life Sci*, 2009, 52: 284–295
- 29 Lin Y, Chen D, Fan Q, *et al.* Characterization of *SoxB2* and *SoxC* genes in *Amphioxus* (*Branchiostoma belcheri*): Implications for their evolutionary conservation. *Sci China Ser C-Life Sci*, 2009, 52: 813–822
- 30 Miao Y, Wu G, Wang L, *et al.* The role of *MC1R* gene in buffalo coat color. *Sci China Life Sci*, 2010, 53: 267–272
- 31 Fu H, Ma H, Zheng C, *et al.* Molecular cloning and differential expression patterns of the regulatory subunit B' gene of PP2A in goldfish, *Carassius auratus*. *Sci China Ser C-Life Sci*, 2009, 52: 724–732
- 32 He J, Irwin M, Zhang Y. *Motilin* and *ghrelin* gene experienced episodic evolution during primitive placental mammal evolution. *Sci China Life Sci*, 2010, 53: 677–682
- 33 Human genome at ten: The sequence explosion. *Nature*, 2010, 464: 670–671
- 34 Hao P, Yu Y, Zhang X, *et al.* The contribution of cis-regulatory elements to head to head gene pairs' co-expression pattern. *Sci China Ser C-Life Sci*, 2009, 52: 74–79
- 35 Zheng L, Qu L. Computational RNAomics: Structure identification and functional prediction of non-coding RNAs *in silico*. *Sci China Life Sci*, 2010, 53: 548–562
- 36 Wang R, Shi L. The evolution of cooperation in asymmetric systems. *Sci China Life Sci*, 2010, 53: 139–149
- 37 Pennisi E. Breakthrough of the year. Human genetic variation. *Sci-*

- ence, 2007, 318: 1842–1843
- 38 Shi J, Wang Y, Huang W. Development and application of genotyping technologies. *Sci China Ser C-Life Sci*, 2009, 52: 17–23
- 39 Wu X, Xiao H. Progress in the detection of human genome structural variations. *Sci China Ser C-Life Sci*, 2009, 52: 560–567
- 40 Tan L, Liu R, Lei S, *et al.* A genome-wide association analysis implicates *SOX6* as a candidate gene for wrist bone mass. *Sci China Life Sci*, 2010, 53: 1065–1072

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.